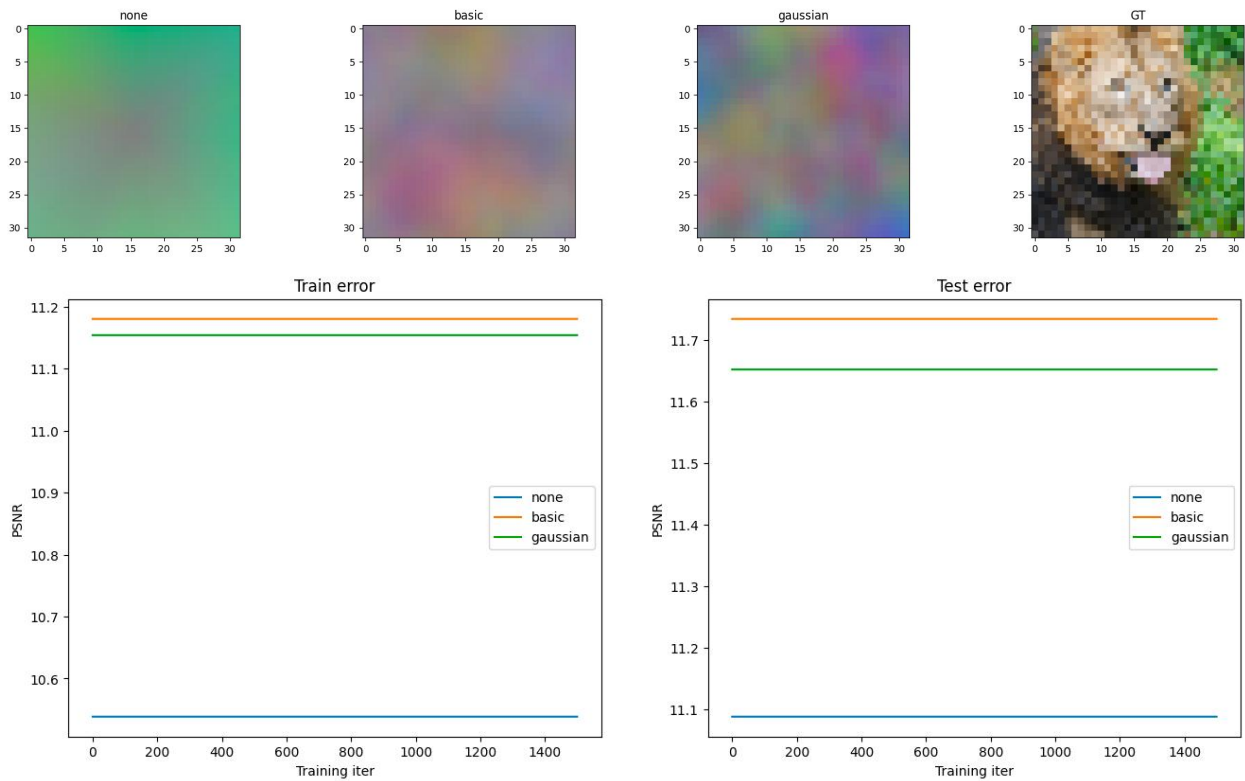# Assignment 2

Name(s): Wenxuan Zhang
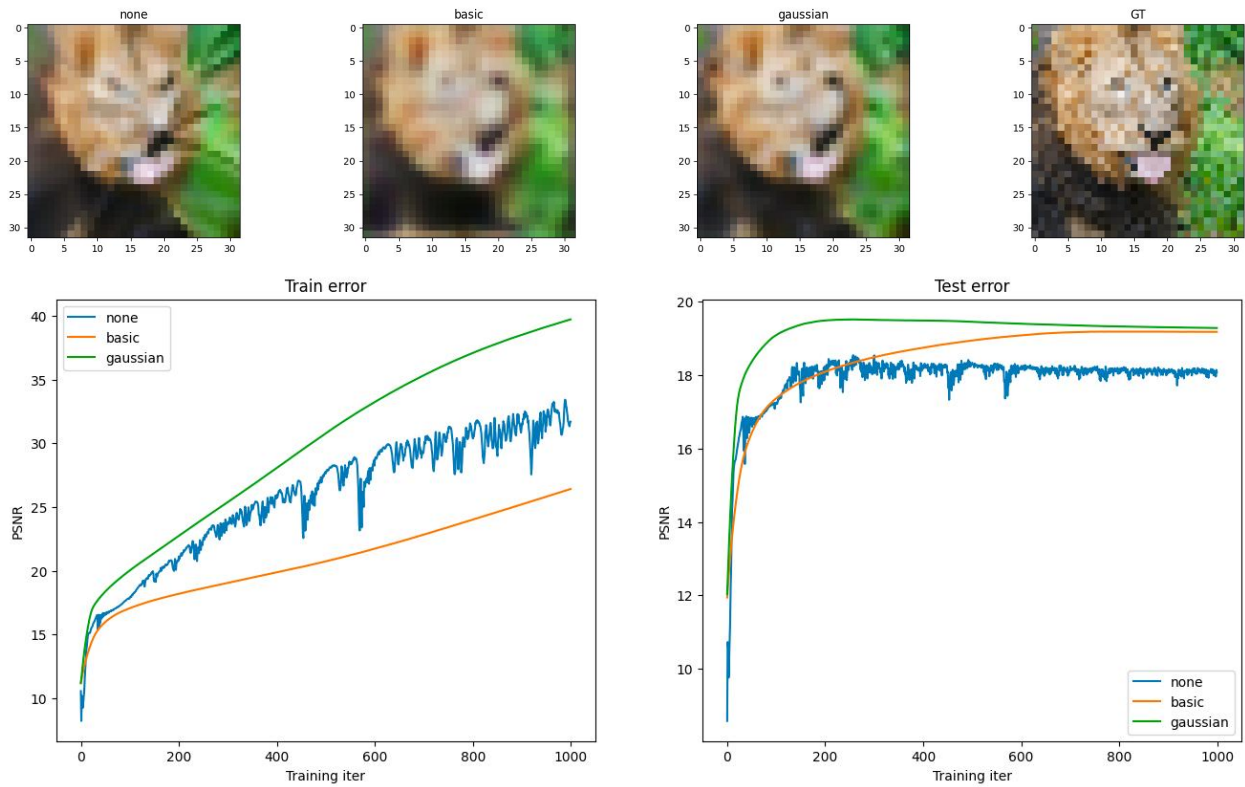## NetID(s): wz47

In each the following parts, you should insert the following:
- Train/test loss plots
- Qualitative outputs for GT, No encoding, Basic Positional Encoding, and Fourier Feature Encoding
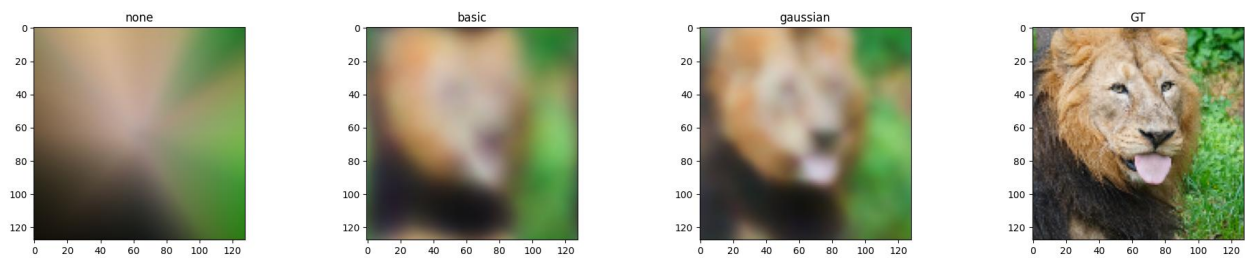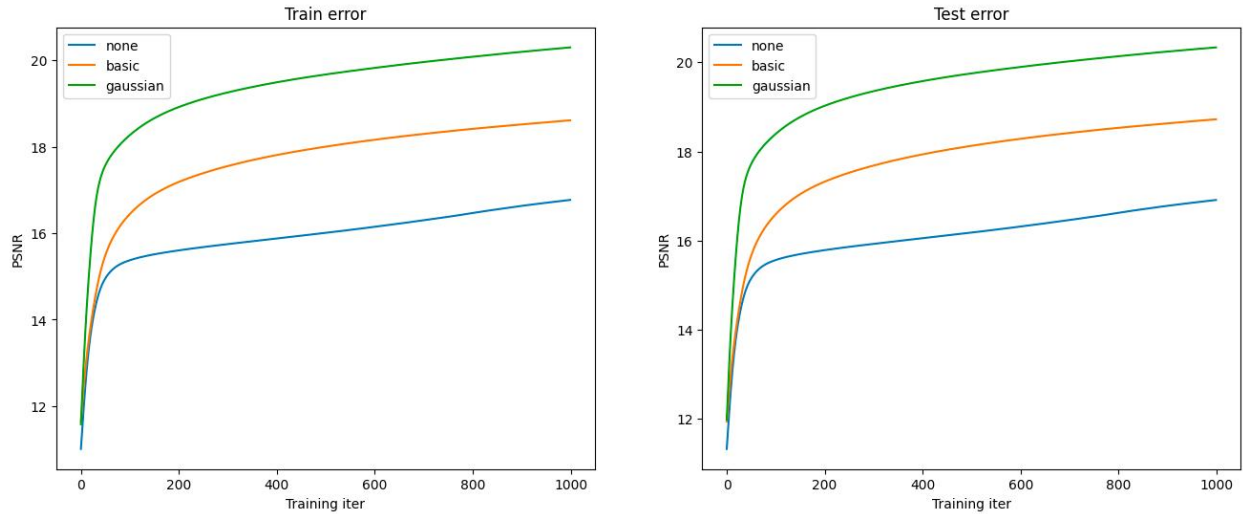
Part 1: Low resolution example - SGD

Part 2: Low resolution example - Adam

Part 3: High resolution example

Part 4: High resolution (image of your choice)

(For this part, you can select an image of your choosing and show the performance of your model with the best hyperparameter settings and mapping functions from Part 3. You do not need to show results for all of the mapping functions.)
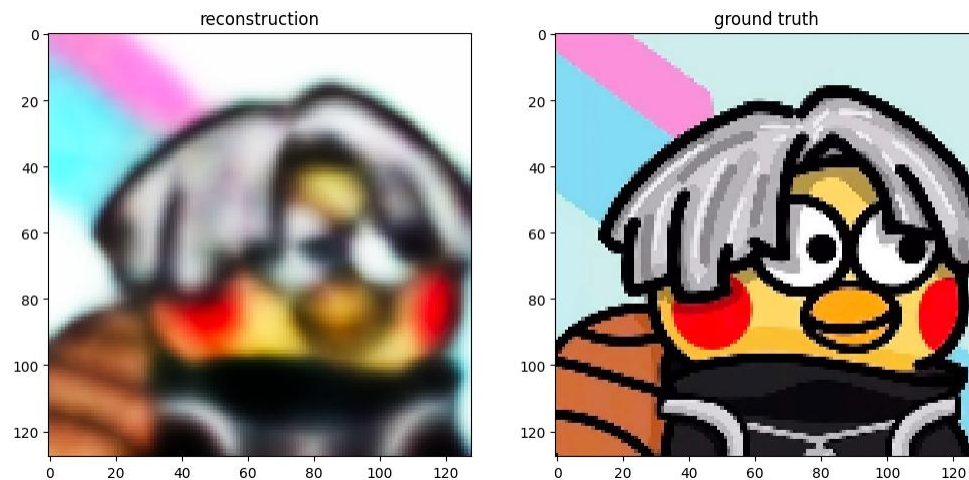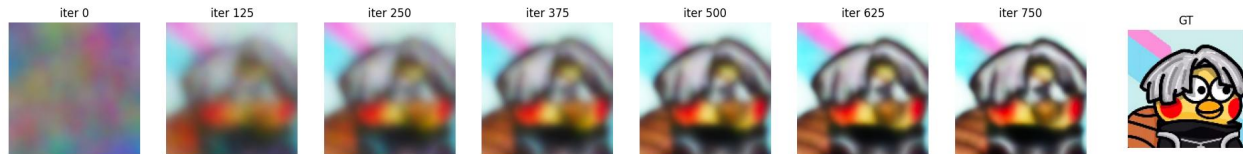
input_size = 512
opt = 'Adam'
Mapping = gaussian
Learning rate = 1e-4
Epoch = 1000

| iter 0 | iter 125 | iter 250 | iter 375 | iter 500 | iter 625 | iter 750 | GT |

Part 5: Discussion

Briefly describe the hyperparameter settings you tried and any interesting implementation choices you made.

For SGD, I set learning rate of [1e-4, 1e-3, 1e-2, 1e-1, 1, 2, 3, 4, 5] to test with three different mapping strategies. It surprisingly turns out that the lower learning rate I set, the more gently the MSE loss tends to grow. More specifically, when using the default learning rate, the MSE becomes a straight line. It turned out that using learning rate of 5, which is obviously too high in most of the deep learning experiment, the MSE can reach the minimum of 0.028337068384071126. I set epoch number to 1500 to match the high learning rate.

How did the performance of SGD and Adam compare?

Adam outperforms SGD in all mapping strategies. For example, in none mapping with learning rate 1e-4, Adam shows MSE for 0.016700557741166015 while SGD only reaches MSE of around 0.06.

How did the different choices for coordinate mappings functions compare?

The gaussian_1.0 mapping tends to be more clear and could reconstruct more structure of ground truth, which might because the randomness in gaussian mapping make the model more rubust to noise and unseen data. The basic mapping is worse than gaussian_1.0 but better than none mapping. The none-mapping strategy shows the least structure of the ground truth.

Do you make any interesting observations from the train and test plots?

The PSNR value grows rapidly in the first 200-400 iterations, and the slows down. Given that PSNR = 20*log(Max_val/sqrt(MSE_val)), the MSE value decreases rapidly in the first 200-400 iterations, which may due to the random initialization of weights.

*What insights did you gain from your own image example (Part 4)?*

1. Choosing ground truth image matters. Because of the image processing issue, I tried more than 10 pictures with different sizes and resolutions, only few of them can present the ground truth in full. With large size, the ground truth might only remains part of the original one. With extremely high resolution, (let's say, 4K), the ground truth resolution is low.
2. Choosing image content matters. I tried different images from city-views to anime and cartoons to animals, and found that the higher the contrast between the subject and the background, the clearer the reconstructed image. The selected one (a cartoon icon) presents the most clear reconstruction image.
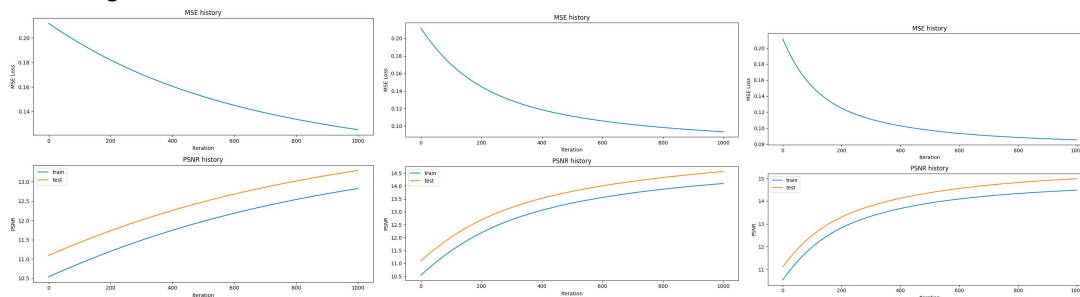
Part 6: Extra Credit

Explain what you experimented with in detail and provide output images.

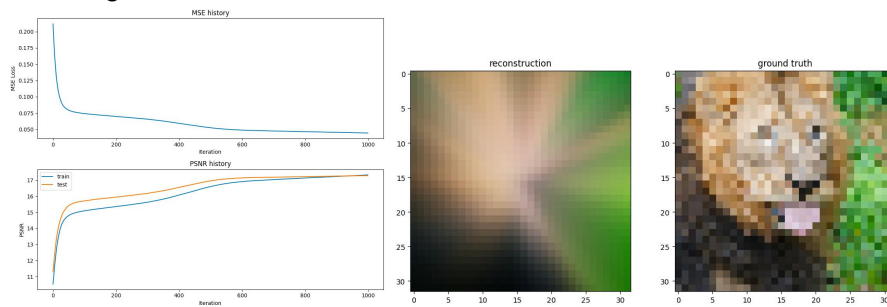1. Testing different learning rate for SGD:
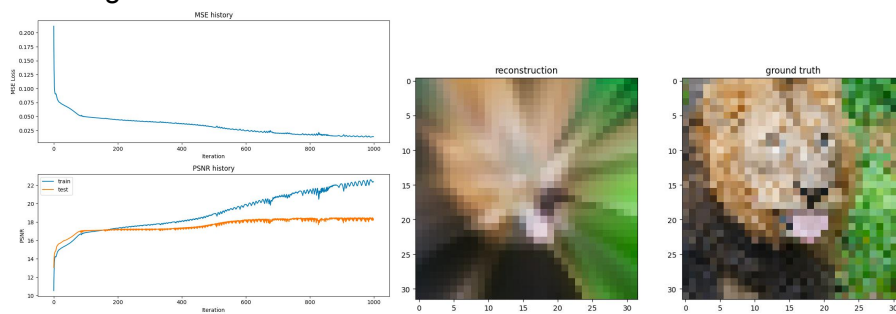Learning rate of 1e-4, 1e-3, 1e-2:
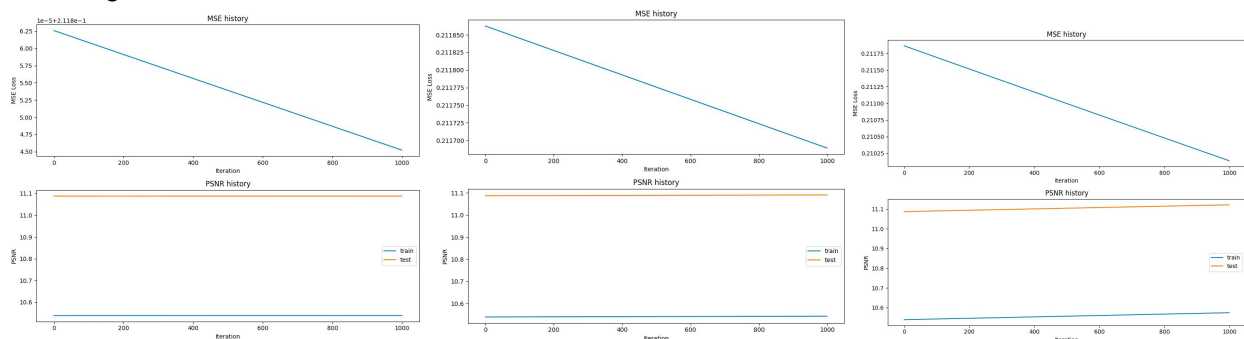
Learning rate of 1, 3, 5:

(The reconstructed images have no obvious differences.
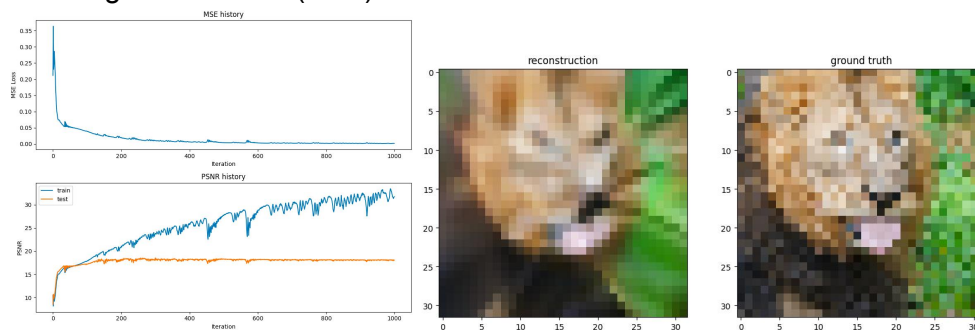
2. Testing different learning rate for Adam:
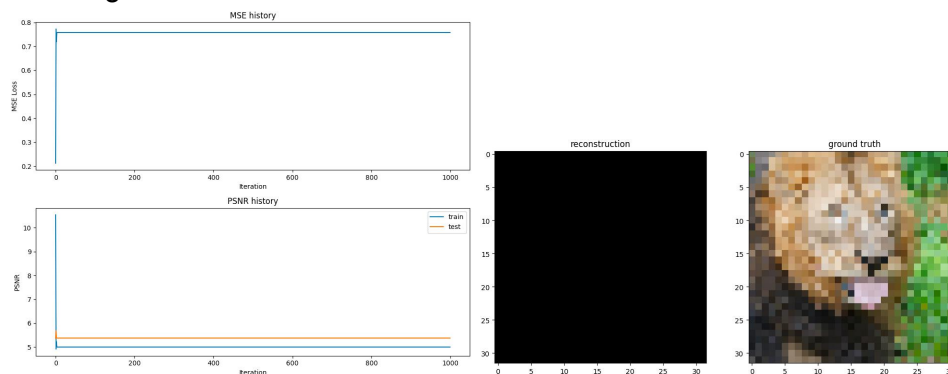
Learning rate of 1e-4:

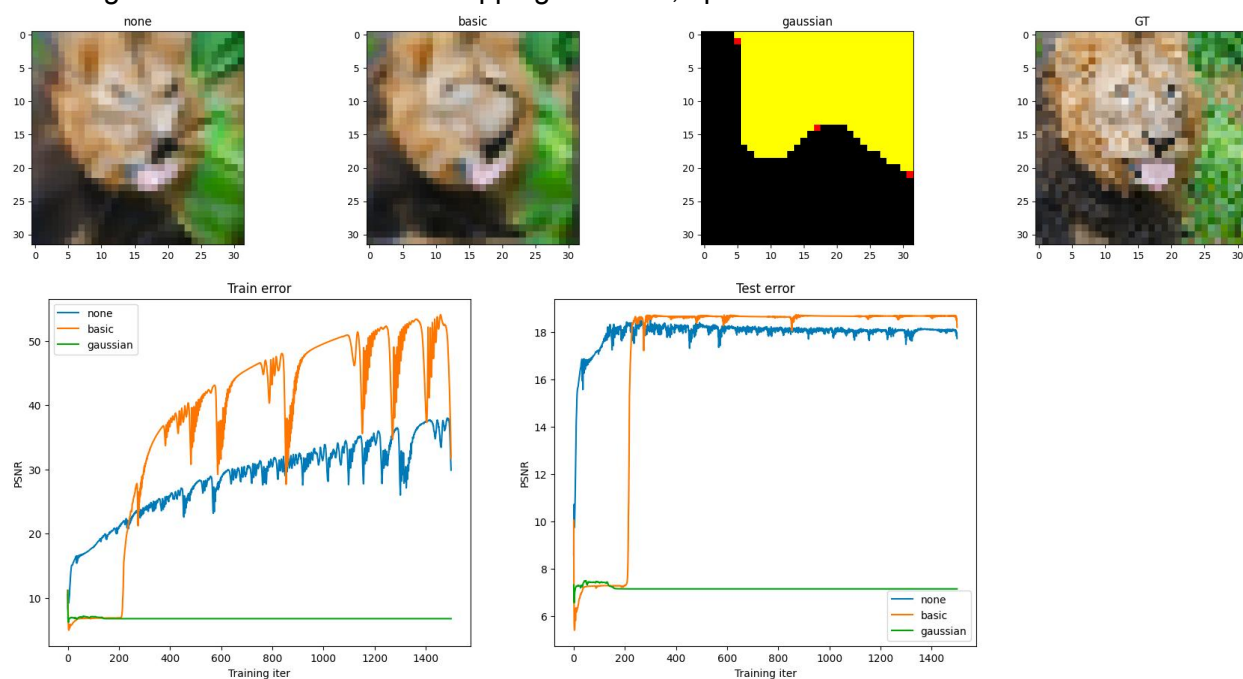### Learning rate of 1e-3:

Learning rate of 1e-2: (Best)

Learning rate of 1e-1:
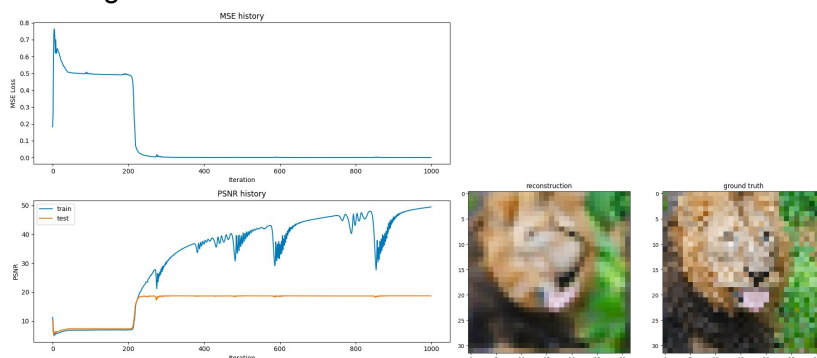
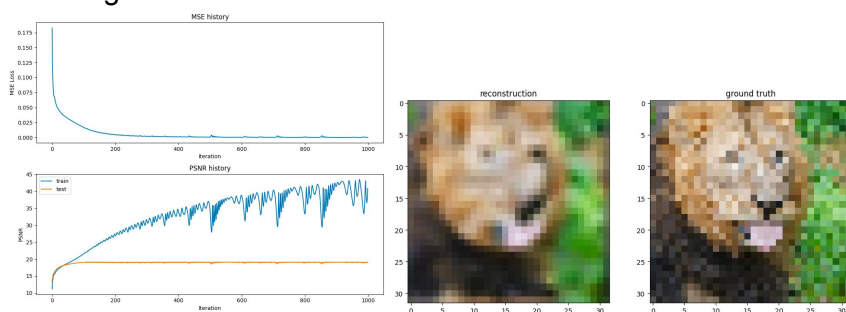Learning rate of 1e-2 of Adam- 3 mapping methods, epoch = 1500:

It turns out 1e-2 is good for none mapping but not suitable for gaussian mapping, so I set different learning rate for different mapping strategy.

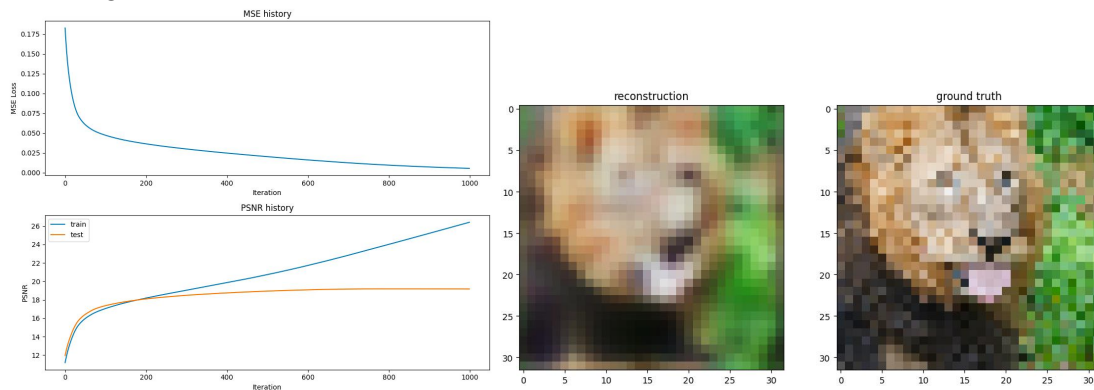3. Different learning rate for Adam - basic:
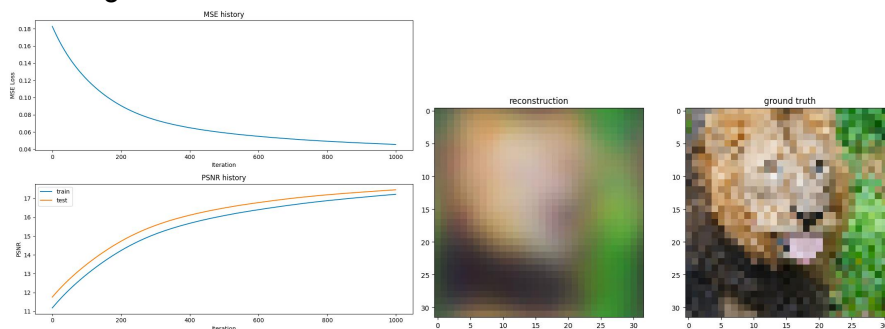Learning rate of 1e-2:

Learning rate of 1e-3:

Final Test MSE 0.011072809903939442

Learning rate of 1e-4:

Final Test MSE 0.010794693637030214

Learning rate of 1e-5:

Final Test MSE 0.016036333613463755

It turns out that learning rate of 1e-4, epoch = 1000 is good for Adam - basic mapping than other rates.
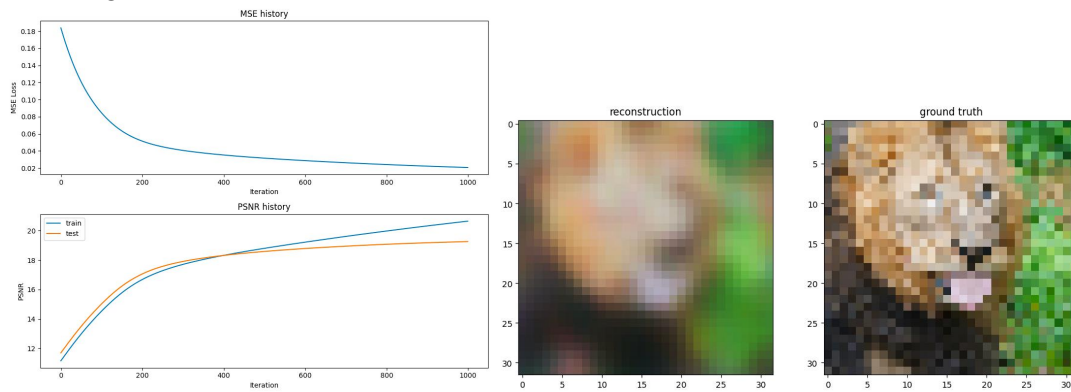
4. Different learning rate for Adam - gaussian
Since in the above experiment, the learning rate of 1e-2(>1e-4, the default learning rate) makes gaussian mapping presents unexpected result, so I decide to experiment with learning rate 1e-4, 1e-5, 1e-6.
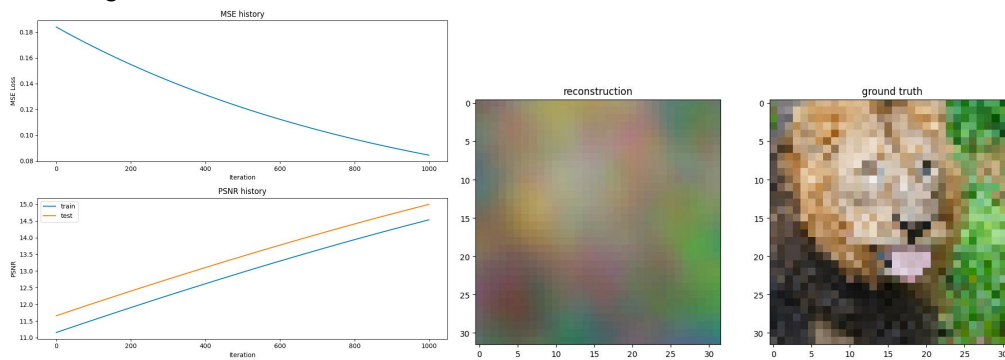
Learning rate of 1e-4:
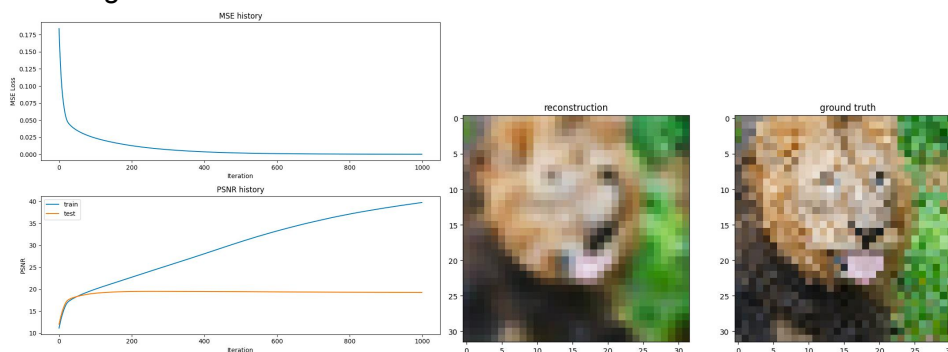
Final Test MSE 0.01053866807796576

Learning rate of 1e-5:

Final Test MSE 0.010591199269930102

Learning rate of 1e-6:

Final Test MSE 0.02827315914386161

Therefore the learning rate of 1e-4 outperforms other learning rate. We could use learning rate of 1e-2, 1e-4, 1e-4 for Adam - non mapping, basic mapping, and gaussian mapping.